In-Network Compute Layer in MobilityFirst Future Internet Architecture (FIA)

Francesco Bronzino, Chao Han, Yang Chen*, Kiran Nagaraja, Ivan Seskar, Xiaowei Yang*, Dipankar Raychaudhuri
WINLAB, Rutgers University and Duke University*

Further information and references are available at: http://mobilityfirst.winlab.rutgers.edu

**MobilityFirst Overview**

- **Layered Names and Fast, Global Name Resolution for Mobility**: All objects including hosts, content, services and abstract context definitions can be assigned network names for direct addressability.
- **Resolution is closely integrated with routing fabric to enable fast and dynamic address bindings**. Name-to-address mappings can be stored/looked up in the order of 10s of ms.
- **Segmented Data Transport with Storage and Edge-Aware Routing**: Data is transported in a hop-by-hop manner leveraging in-network storage and information about edge network state to address variability in access/edge networks, particularly wireless.

**Cloudlets: Compute Layer Service Deployment in ISP networks**

Cloudlets - compute hosting platforms closely integrated with network fabric - may be deployed at one or more Points of Presence (PoPs) and coordinated by a Domain Controller run by the ISP.

**Compute in Contextual and Sensing Cases**

- **In-network Processing and Aggregation of Sensor data**: Sensed data from vehicles and other in-field sensors can be aggregated in the network by a compute layer service explicitly requested by the originator of the sensed data, thus reducing load on a centralized server.
- **Dynamic binding for Context GUID**: A local context defined as ‘unoccupied cabs in location X’ can be named using a GUID and resolved dynamically by an in-network compute layer service. The service may pull information from a web-based dispatch service to determine what end points qualify for delivering a request message addressed to this GUID.

**Compute Layer Opportunities**

- Provides easy extensibility/upgrade options for data plane
- ISPs can use in-network computing to provide value added services such as caching, security, and contextual services
- ISPs can monetize their in-network cloud deployments by offering a Platform as a Service (PaaS) solutions to application service providers
- Application and content providers can deploy close-to-client solutions to minimize access latency

**Data Packet Handling for Compute**

- **Explicit request for compute layer processing with Service type and custom Extension Header**, Extension header encodes GUID of service and any arguments besides the payload.
- For instance, to request a transcoding operation on the payload, extension header contains at the minimum the transcoding service’s GUID and a target bitrate
- Router redirects data packets with the above set in the headers to a co-located compute service registered a priori

**In-Network Transcoding using MF Compute Layer**

1. P1 publishes content C (matched over network)
2. Mobile M requests video C
3. Lookup of C resolves recursively to Network 19
4. Video Server streams video segments with compute header for Transcoder T
5. Lookup of M resolves to Network 53a
6. Router with compute layer plugin forwards chunk to T
7. T transcodes content to target bitrate suitable for M

**GENI Demo of MF Compute Layer**

- Multiple sites with InstaGENI, some with WMIPv6: Rutgers*, Wisconsin*, NYU/NYU-Poly*, Utah, NYSERNET, GPO, UIUC. Multipoint VLAN connects sites to enable layer-2 connectivity for non-IP MF protocol
- DASH video stream server and VLC player client. Video segment requests by client are translated to MF content get requests. Server includes a compute extension header in data packets so high quality content may optionally be transcoded to a rate suitable for client
- Transcoder invoked by co-located router if the monitored client access bandwidth drops below original encoded video rate